

PB87-193629

Resampling and Extreme Value Statistics in
Air Quality Model Performance Evaluation

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

1985



TECHNICAL REPORT DATA

(Please read Instructions on the reverse before comp)

1. REPORT NO. EPA/600/J-85/486		2.		PB87-193629	
4. TITLE AND SUBTITLE Resampling and Extreme Value Statistics In Air Quality Model Performance Evaluation (Journal Version)				5. REPORT DATE 1985	
				6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) S.T.Rao, G. Sistla, V. Pagnotti, W.B. Petersen, J.S. Irwin, D.B. Turner				8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Division of Air New York State Department of Environmental Conservation Albany, N.Y. 12233				10. PROGRAM ELEMENT NO. CDTA1D/04- 3012 (FY-85)	
				11. CONTRACT/GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Sciences Research Laboratory--RTP, NC Office of Research and Development U. S. Environmental Protection Agency Research Triangle Park, NC 27711				13. TYPE OF REPORT AND PERIOD COVERED	
				14. SPONSORING AGENCY CODE EPA/600/09	
15. SUPPLEMENTARY NOTES REF: Atmos. Environ., 19(9):1503-1518, September 1985					
16. ABSTRACT Specific data analysis techniques that will reveal the performance of air quality models in simulating the measured concentrations' cumulative distribution are discussed in this paper. The paper presents two types of analysis to compare model predictions with the measurements. In one analysis, extreme value statistics and the fitting of tail exponential distributions to both measured and predicted values are used in various ways to see if the measured and predicted values fit such distributions and to what degree the higher values of the cumulative frequency distributions coincide. In the second analysis, a resampling (bootstrap technique is used to develop non-parametric confidence intervals for the entire cumulative distribution of the measured concentrations, and to derive empirical distributions for central tendency statistics and for differences between measured and predicted mean and median values. The analysis is focused so as to show (1) why the resampling is necessary and the degree to which mistaken judgements can be made with and without the technique, and (2) comparisons between the discriminating capabilities of "tail fit" type model evaluation and one using the resampling technique. It is shown that both the bootstrap and extreme value statistics are needed to quantify the uncertainty associated with the model predictions.					
17. KEY WORDS AND DOCUMENT ANALYSIS					
a. DESCRIPTORS		b. IDENTIFIERS/OPEN ENDED TERMS		c. COSATI Field/Group	
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC		19. SECURITY CLASS (This Report) UNCLASSIFIED		21. NO. OF PAGES 17	
		20. SECURITY CLASS (This page)		22. PRICE	

RESAMPLING AND EXTREME VALUE STATISTICS IN AIR QUALITY MODEL PERFORMANCE EVALUATION

S. T. RAO, G. SISTLA, V. PAGNOTTI

Division of Air Resources, New York State Department of Environmental Conservation, Albany, NY 12233,
U.S.A.

and

W. B. PETERSEN*, J. S. IRWIN* and D. B. TURNER*

Meteorology and Assessment Division, U.S. Environmental Protection Agency, Research Triangle Park,
N.C. 27711, U.S.A.

(First received 26 September 1984 and in final form 28 January 1985)

Abstract—Specific data analysis techniques that will reveal the performance of air quality models in simulating the measured concentrations' cumulative distribution are discussed in this paper. The paper presents two types of analysis to compare model predictions with the measurements. In one analysis, extreme value statistics and the fitting of tail exponential distributions to both measured and predicted values are used in various ways to see if the measured and predicted values fit such distributions and to what degree the higher values of the cumulative frequency distributions coincide. In the second analysis, a resampling (bootstrap) technique is used to develop non-parametric confidence intervals for the entire cumulative distribution of the measured concentrations, and to derive empirical distributions for central tendency statistics and for differences between measured and predicted mean and median values. The analysis is focused so as to show (1) why the resampling is necessary and the degree to which mistaken judgements can be made with and without the technique and (2) comparisons between the discriminating capabilities of 'tail fit' type model evaluation and one using the resampling technique. It is shown that both bootstrap and extreme value statistics are needed to quantify the uncertainty associated with the model predictions. These techniques are applied to the model predictions from RAM and measurements from the Regional Air Pollution Study data base to demonstrate their utility in the quantitative assessment of the model performance.

Key word index: Air quality models, model evaluation, resampling, bootstrap, extreme-value theory, statistical analysis, order statistics.

1. INTRODUCTION

In recent years, considerable attention has been given to the model performance evaluation and quantification of modeling uncertainty. Under a cooperative agreement with the U.S. Environmental Protection Agency (EPA), the American Meteorological Society (AMS) assembled a team of experts and conducted two workshops to suggest statistical techniques that would reveal the performance of air quality models. The specific recommendations of the workshop members have been summarized by Fox (1981, 1984). Most of the existing literature on model evaluation and the AMS recommendations address the problem of assessing the overall performance or central tendencies of model behavior. Little attention has been paid to the development and application of statistical tools to evaluate the model performance in predicting the upper tail of the concentrations' cumulative distribution even though this is the region of primary concern to the decision makers in the regulatory agencies.

The object of this investigation is to improve upon the previously employed methodologies through the utility of some specific statistical analysis techniques in the quantitative evaluation of the performance of models. To this end, extreme value statistics are employed to determine the model's capability in simulating the tails of the observed concentrations' cumulative distribution. Also, the resampling or bootstrap technique is applied to develop non-parametric confidence intervals for each percentile value of the concentrations distribution. These techniques enable the assessment of the model's behavior both on a paired (observed and predicted value at the same time and location is considered as a pair) and unpaired basis. In this way, model's performance from the scientific and regulatory or operational viewpoints can be evaluated.

In this paper, methodologies and rationale for the application of resampling and extreme value analyses are provided in Section 3. The confidence intervals determined from the known properties of the full parent distribution are compared with those determined from the bootstrap and order statistics in Section 4. It is shown that the extreme probability

*On Assignment from NOAA, U.S. Dept. of Commerce.

points determined from the full parent distribution may exhibit substantial irregularity even when the underlying distribution is known *a priori*. The resampling and extreme value statistics to quantitatively assess the performance of air quality models are also discussed in Section 4. The utility of these techniques is demonstrated with the measured and model predicted SO₂ concentrations in the St. Louis area. The results are summarized in Section 5.

2. DATA BASE FOR MODEL EVALUATION

Measured hourly SO₂ concentrations from the Regional Air Pollution Study (RAPS) data base at one of the 13 locations in St. Louis, MO and corresponding hourly SO₂ concentrations predicted by RAM, an EPA air quality model, are used to demonstrate the utility of resampling and extreme value statistics in air quality model performance evaluation. A description of the RAPS data base can be found in Schiermeier (1978). The details on the RAM formulation can be found in Turner and Novak (1978). The SO₂ concentration measurements were subjected to extensive screening, and procedures for checking the completeness of the sample collection and the quality of the measurements in the RAPS data base were developed (Irwin, pers. comm. 1983).

A subset of these data having a sampling collection of greater than 54 minutes in the hour, and having a span and zero drift during the hour less than 5 and 0.5%, respectively, was collected in this study. Hourly SO₂ measured concentrations and their corresponding model predicted concentrations for similar atmospheric stability and wind speed conditions at site 108 of the RAPS network were assembled. Thus, the data for four atmospheric stability categories (unstable, day neutral, night neutral and stable) and three wind speed classes (< 2 m s⁻¹, 2–5 m s⁻¹, and > 5 m s⁻¹) at this site are analyzed.

3. DATA ANALYSIS METHODS

The statistical techniques employed in this study are briefly summarized below.

3.1. Resampling analysis

Recently, computer-intensive statistical methods have been developed to make accurate and versatile statistical inferences. The bootstrap and jackknife techniques are resampling methods which provide an estimate of the uncertainty or statistical accuracy of the parameter of interest without requiring any distributional assumptions regarding the data. The mathematical framework and the details on the bootstrap technique can be found in Efron (1982). Briefly,

(a) Let $X_i = x_i$, $i = 1, 2, \dots, n$ be the data points in the original sample where the X_i are independent and identically distributed (iid) according to some unknown probability distribution F .

(b) Draw n new points (X_i^*) independently and with replacement from the original sample so that each new point is an independent random selection of one of the n original data points. These new points are called the 'bootstrap sample'. That is, in each bootstrap replication, an item (a data point) is selected randomly from the original sample and noted as an entry in the

bootstrap sample, and returned to the original sample. Thus, some of the original points will have been selected zero times, some once, some twice, or more. The expected proportion of points in the original sample absent from the bootstrap sample is $(1 - 1/n)^n$.

(c) This gives a bootstrap empirical distribution function $\hat{F}^*$ of the n new values.

(d) Determine the parameter of interest $\hat{\theta}^*$, which may be the mean, median, or a certain percentile value, from $\hat{F}^*$.

(e) Repeat steps (b)–(d), independently, a large number of times, say N , giving bootstrap values for the parameter $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*j}, \dots, \hat{\theta}^{*N}$.

(f) The bootstrap standard deviation (σ_{boot}) is an estimate of the spread in the distribution $\hat{F}$. If the true F equalled $\hat{F}$, then σ_{boot} is also an estimate of the spread in F .

Let $[a^*, b^*]$ be the central 68% interval for the $\hat{\theta}^*$ values, i.e.

$$\# \frac{[\hat{\theta}^{*j} < a^*]}{N} = 0.16; \quad \# \frac{[\hat{\theta}^{*j} < b^*]}{N} = 0.84.$$

$$\text{Then } \sigma_{\text{boot}} = (b^* - a^*)/2.$$

If we could let the number of bootstrap replications $N \rightarrow \infty$, then the spread σ_{boot} would exactly equal σ , the spread in the true distribution F . In many applications, a few hundred bootstrap replications are sufficient to determine the parameter of interest and its confidence intervals with a reasonable accuracy (see Efron, 1982). Some of the limitations associated with the bootstrap method are as follows.

(i) The data in the original sample should be a random selection of items from a homogeneous population. That is, the data are statistically independent of one another. Because of the nature of the resampling procedure, any correlation or temporal/spatial dependence in the original sample will not be maintained in the bootstrap sample. For practical purposes, it is sufficient that the time-series data being subjected to bootstrap analysis consist of data values sufficiently far apart in time so that they can be considered statistically independent.

(ii) Sample maxima (minima) greater (smaller) than those present in the original sample cannot be produced by the bootstrap technique. However, the bootstrap method can be modified by parametrically smoothing and extending the tail of each bootstrap sample to avoid the limitations associated with the sample maxima or minima.

If $d = (O - P)$, where O is the observed and P is the corresponding predicted concentration, the construction of $F(d)$ from the bootstrap technique outlined above will enable the estimation of model uncertainty. Further, it is possible to develop non-parametric confidence intervals based on the empirical distributions from the resampling technique for the entire concentration cumulative distribution, and whether air quality model predictions lie within the confidence intervals for the measured values at each probability

point can be determined to assess the simulation capability of the model.

3.2. Extreme value statistics

The model's capability in predicting the extreme or design values is of primary concern to the decision makers in the regulatory agencies. The knowledge of the underlying distribution for the entire range of data, the parent distribution, is required for describing the statistical behavior of all data values. Consideration of the upper tail of the parent distribution simplifies the problem since the tail distribution is easier to model statistically. Further, little information is lost by discarding the smaller values since they offer little insight into the behavior of the tail of the underlying distribution (Galfond, 1982). Hirtzel (1984) examined the distribution of the maximum value and its parameters, and selected moments (i.e. the mean and variance of the maximum) for the cases where the initial distribution is normal, gamma or log-normal, and found that the estimates of these parameters and moments are highly dependent upon the type of the initial distribution assumed. The direction or trend (i.e. increase or decrease with increasing sample size) as well as the magnitude of a particular parameter or moment may be quite different. Hence, the use of the parent or underlying distribution for the entire range of the data to predict the extreme values could lead to erroneous results regarding the maxima. In the case of air pollution data, evidence exists that the data are sometimes log-normally distributed, and other times or places a gamma distribution is observed. One could obtain a good fit to a theoretical distribution in the central portion of the data, but fail to achieve a good fit in the tails. Thus, the use of the parent distribution describing the entire data set may not be appropriate to assess the behavior of the extremes. Therefore, in this study, attention is also paid to the characterization of the tail distributions.

Either the exact or the asymptotic theory of extreme values can be utilized to assess the model behavior in simulating the extreme values. Of course, as in the resampling theory, the data should satisfy the iid assumption. The major limitation of the exact theory of extreme values is that the underlying distribution of the data in the selected upper percentile should be known *a priori*. If the distribution, $F(x)$, is known, then the cumulative probability distributions for the first, second, or the m th order statistic can be developed from the relation

$$G_{m,n} = \binom{n}{m} m \int_0^{F(x)} p^{n-m} (1-p)^{m-1} dp, \quad (1)$$

where $G_{m,n}$ is the cumulative distribution function (CDF) for the m th highest value for a sample size of n , $F(x)$ is the distribution of the n original values, and P is the probability of any value x less than the maximum x_m , over the n values.

There is strong evidence that the tails of most parent distributions can be described by an exponential distribution (Breiman *et al.*, 1979, 1981; Galfond, 1982) having the form

$$F(x) = 1 - e^{-\lambda(x-B)} \quad \text{for } x > B; \lambda > 0, \quad (2)$$

where λ and B are the shape and location parameters. The parameters characterizing this distribution can be determined using curve-fitting techniques, such as maximum likelihood estimation, the method of moments or linear least-squares fit (David, 1981). The distribution fitted to the original data can be accepted or rejected using non-parametric statistical tests. As indicated before, the results using the exact theory of extreme values are strongly dependent upon the knowledge of the functional form of $F(x)$, the distribution of the pollutant concentration data. Even though there is strong evidence that the data in the upper tail of the original sample are describable by the exponential distribution, approaches that are less sensitive to the nature of the underlying distribution deserve consideration.

For a wide class of initial or parent distributions, the m th extreme distributions approach a common distribution in the asymptotic limit, for large n . Irrespective of the initial distribution function $F(x)$, there are only three types of extreme value asymptotic forms (see Tabony, 1983). Again, the random variable being subjected to the asymptotic extreme value analysis should satisfy the condition that the variable is statistically independent and identically distributed. If the initial distribution is of the exponential family, including normal, log-normal, Weibull, Gamma and related distributions, then the asymptotic extreme value distribution functions (of the Type I) of the first (Y_1) and second (Y_2) maxima as the sample size approaches infinity are

$$G_1(Y_1) = \exp[-\exp(-Y)] \quad (3)$$

$$G_2(Y_2) = [1 + \exp(-Y_1)]G_1(Y_1);$$

$$\text{if } Y_2(x) = Y_1(x) + \ln 2 \quad (4)$$

under the conditions $-\infty < Y < +\infty$; and $Y = A(x - U)$, where Y is the reduced variate, x is the value of the random variable, U is the mode of the extreme value distribution and A is the inverse of the Gumbel slope (see Gumbel, 1962). Once the extreme value distribution is determined, the 95% confidence limits for the distribution can be identified.

4. RESULTS AND DISCUSSION

4.1. Analysis with hypothetical data from a theoretical distribution

Bootstrap replication is a method to generate new samples randomly from the original sample. Based on this technique, non-parametric confidence intervals for the parameters of interest can be developed. Extreme value statistics enable the documentation of the be-

havior of the extreme values. If the parent distribution is known *a priori*, one can estimate the maximum and determine its confidence intervals. An exercise is performed here with a known theoretical distribution using normal probability theory and the bootstrap technique. Then, a comparison is made of results obtained from the bootstrap as well as the exact and asymptotic extreme value analyses of the tail of this parent distribution with those derived from the known properties of the parent distribution.

4.1.1. *Data base for the test case.* In order to compare the confidence intervals for the selected percentiles and the maximum value determined from the knowledge of the parent distribution with those determined from either the bootstrap or the extreme value analysis, a theoretical distribution of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-Z^2/2] \quad (5)$$

is employed, where $Z = (x - \bar{x})/\sigma$, $\bar{x}$ is the mean and σ is the standard deviation. The transformation or change of variable Z is in common use, since it requires only one table for all normal distributions regardless of their different means and variances. The distribution of Z is normal with zero mean and unit standard deviation. The CDF of the normal distribution is given by

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp[-(x - \bar{x})^2/2\sigma^2] dx. \quad (6)$$

Based on the normal probability theory, 1000 values of the variable x are generated, and their cumulative probability distribution is presented in Fig. 1(a). These values belong to the normal distribution $N(3.4, 1.63)$ with a mean ($\bar{x}$) of 3.4, and the standard deviation (σ) of 1.63. The 95% confidence limits for the fitted line (developed using $y = \bar{L} \pm k_1 S_y$, where $\bar{L}$ is the observed value of the line at each percentage point, k_1 is the multiplier dependent upon the sample size, and S_y is the standard error of estimate of the line), presented in Fig. 1(a), confirm that the variable is normally distributed.

4.1.2. *Analysis with the bootstrap method.* The 1000 original data points representing the normal population are subjected to the resampling procedure, and the mean and standard deviation for the bootstrap sample in each bootstrap replication are computed. In each replication, the numerical values for selected percentiles (corresponding to specific Z values) are determined from the normal probability theory, and the mean and the standard deviation of the selected percentile values over the 1000 bootstrap replications are then determined as follows

$$\text{Mean: } \hat{\theta}^{**} = \sum_{j=1}^N \hat{\theta}^{*j}/N, \quad (7)$$

standard deviation:

$$\hat{\sigma} = \left\{ \frac{1}{N-1} \sum_{j=1}^N [\hat{\theta}^{*j} - \hat{\theta}^{**}]^2 \right\}^{1/2}.$$

Thus, this resampling procedure represents a parametric smoothing of the bootstrap technique, or an example of bootstrap modification. The confidence interval for the percentile value of interest is determined from the expressions $\hat{\theta} \pm C_\alpha$, where C_α is the number taken from the normal or t -tables. The results of these analyses are presented in column 3 of Table 1.

Since the bootstrap method cannot provide values larger or smaller than those present in the original sample, normal probability theory is used in determining the lower and upper bounds at the 95% level of significance for upper percentile values rather than the numerical values at each percentile provided by the resampling procedure. Of course, the underlying assumption is that the data in each bootstrap replication belong to the normal distribution. Since the original sample is normally distributed, it is assumed that the bootstrap estimate of the central tendencies (mean and standard deviation) in each replication can be used to determine the values corresponding to the tail probabilities from the normal theory. The confidence intervals estimated from the modified bootstrap method using Equation (7) can be compared with the 95% confidence intervals for the original sample based on normal probability theory in Table 1 (see columns 2 and 3). In general, the confidence bounds determined from the modified bootstrap analysis are smaller than the 95% confidence bounds based on the normal distribution for the original sample. It should be noted that the value at the 99.9% probability point determined from the normal distribution to the original sample is 8.42 even though the maximum value, occurring at this probability, in the original sample is 8.1. This suggests that even the plot for a sample of 1000 drawn from a population known to be normal exhibits substantial irregularity at the tails. Thus, the use of the full parent distribution or the parametric smoothing of the bootstrap method as applied here could yield values for the tail probabilities that may be different from the original sample.

As indicated above, a parametric smoothing procedure for the bootstrap method is used to remove the inherent constraint that the bootstrap technique cannot generate values outside the range in the original sample. Having observed that the modified bootstrap provides a narrower 95% confidence interval for the tail, an analysis is performed here using the bootstrap method on the original sample without any modification to the method. The original sample is again subjected to 1000 bootstrap replications and the values corresponding to the selected percentiles are determined through rank-ordering the data in each bootstrap replication. The values at each upper percentile from 1000 replications are used to determine the mean value at each probability point. The percentile method (Efron, 1982) is used to determine the approximate 95% confidence intervals for the parameters of interest. Briefly, let

$$\text{CDF}(t) = \text{PROB}_\bullet[\hat{\theta} \leq t] \quad (8)$$

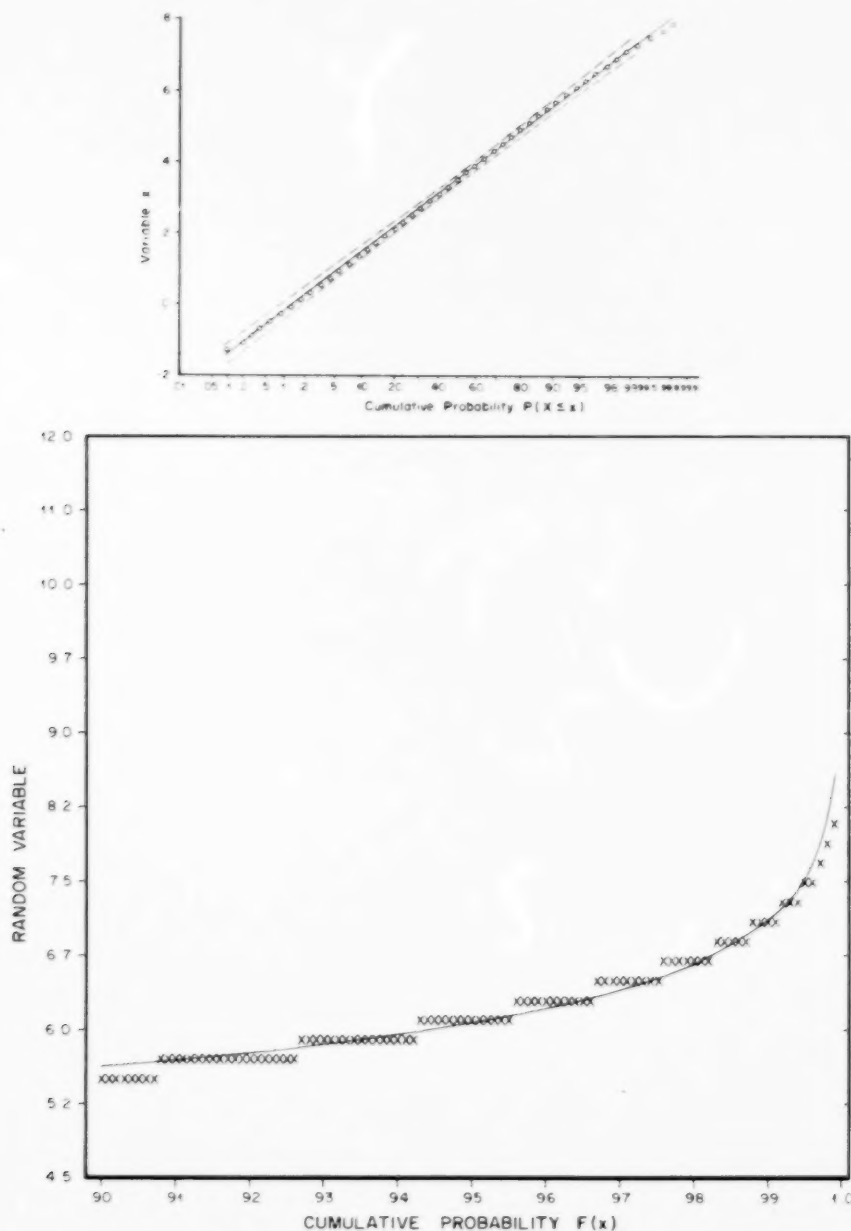


Fig. 1. (a) The cumulative probability distribution of the 1000 values chosen from a normal distribution. The 95% confidence intervals for the distribution based on the normal probability theory are also presented. (b) Exponential fit to the upper 10% of the initial distribution. The individual values (x 's) in the tail of the initial distribution are shown to provide a visual examination of the exponential fit through the tail probability points.

be the CDF of the bootstrap distribution of $\hat{\theta}^*$. For a given α between 0 and 0.5, define

$$\hat{\theta}_{\text{low}}(\alpha) = \hat{C}\hat{D}F^{-1}(\alpha); \quad \hat{\theta}_{\text{up}} = \hat{C}\hat{D}F^{-1}(1 - \alpha) \quad (9)$$

usually denoted simply $\hat{\theta}_{\text{low}}$, $\hat{\theta}_{\text{up}}$. The percentile method consists of taking

$$[\hat{\theta}_{\text{low}}(\alpha), \hat{\theta}_{\text{up}}(\alpha)] \quad (10)$$

as the approximate $[1 - 2\alpha]$ central confidence interval for θ . Since

$$\alpha = \hat{C}\hat{D}F(\hat{\theta}_{\text{low}}), \quad (1 - \alpha) = \hat{C}\hat{D}F(\hat{\theta}_{\text{up}}) \quad (11)$$

Table 1. Comparisons between the parametric and non-parametric estimates of confidence intervals for selected percentile values of a normally distributed variable with a mean $\bar{x}$ and a standard deviation σ . A sample size of 1000 values is chosen here representing a normal distribution $N(3.4, 1.63)$

Percentile	From the parent distribution $N(3.4, 1.63)$		From the parametric smoothing of the bootstrap method		From the bootstrap method		From the order statistics	
	Value of the random variable (x)	95% confidence interval	Value of the random variable (x)	95% confidence interval	Mean value of the random variable (x)	95% confidence interval	Lower bound	Upper bound
90th	5.49	± 0.16	5.49	± 0.13	5.51	5.3	5.32	5.67
91	5.58	± 0.17	5.59	± 0.13	5.62	5.5	5.41	5.77
92	5.69	± 0.17	5.69	± 0.14	5.71	5.5	5.51	5.88
93	5.80	± 0.18	5.81	± 0.14	5.83	5.7	5.62	6.01
94	5.92	± 0.19	5.93	± 0.14	5.95	5.7	5.74	6.14
95	6.08	± 0.19	6.08	± 0.15	6.09	5.9	5.88	6.31
96	6.25	± 0.20	6.25	± 0.15	6.26	6.1	6.04	6.50
97	6.46	± 0.22	6.46	± 0.16	6.44	6.3	6.23	6.74
98	6.74	± 0.23	6.76	± 0.17	6.69	6.5	6.48	7.07
99	7.17	± 0.25	7.19	± 0.18	7.06	6.7	6.86	7.62
99.9 (Highest)	8.42	± 0.33	8.43	± 0.23	7.99	7.5	7.77	9.95
Highest* (99% probability)	8.18	± 1.09						

*Based on the 95% confidence interval for the best-fit line on the extreme value probability chart using the double exponential or Gumbel (Type I) extreme value distribution.

†The 95% confidence intervals are based on the percentile method to the bootstrap distribution at each probability point.

the percentile method interval consists of the central $[1 - 2\alpha]$ proportion of the bootstrap distribution.

The means over the 1000 replications, and the approximate 95% confidence intervals [using Equation (11)] for each selected upper percentile based on the bootstrap method are presented in column 4 of Table 1. The mean values at each probability point as well as the 95% confidence intervals for the quantiles based on the direct application of the bootstrap method do not differ noticeably from those determined by the normal distribution to the original sample. The 95% confidence intervals about the mean values of the quantiles based on normal theory and the approximate 95% confidence interval developed from the percentile method to bootstrap distribution are seen to be in excellent agreement. Since the confidence intervals, based on the percentile method, computed from the bootstrap technique do not require any assumptions regarding the underlying distribution, the application of bootstrap technique provides robust estimates of the confidence intervals for the parameters of interest.

4.1.3. Comparison between the confidence intervals derived from the order statistics and the bootstrap technique. The 95% confidence intervals for each percentile value are also determined from the order statistics by substituting Equation (6) into Equation (1), and solving Equation (1) for selected m th highest values for the sample size (n) of 1000 values. The confidence intervals determined from the order statistics approach are presented in column 5 of Table 1. Again, excellent agreement between the confidence bounds estimated from the order statistics and those determined from the direct application of bootstrap method is evident up to the 97th percentile (see columns 4 and 5 in Table 1). Whereas the second and third columns in Table 1 represent the confidence intervals about the mean value at each quantile, the last two columns in Table 1 reflect the range of outcomes from the order statistics. Beyond the 97% probability point, the results from the bootstrap method deviate significantly from those determined from the order statistics. The reason for this discrepancy is the fact that there are insufficient samples at these quantiles to perform meaningful resampling. It seems that at least 30 data points are necessary to perform the resampling analysis.

The analysis with the hypothetical data known to be normally distributed indicates that the numerical value for the variable corresponding to the extreme probabilities determined from the full parent distribution may exhibit substantial irregularity. Although the bootstrap method provides non-parametric confidence intervals at each probability point, it should be noted that the upper and lower bounds for the extreme probability points are constrained since the method cannot generate values outside the range of the values present in the original sample, and there is also a lack of sufficient number of samples at extreme probability points from which a meaningful resamp-

ling can be made. Thus, the bootstrap method or the full parent distribution cannot provide reliable information on the tail behavior. Also, this analysis suggests that bootstrap as modified here is no panacea for the limitations of this approach to describe the tails.

4.1.4. Analysis with the exact theory for extreme values. Having seen the limitations to the use of the full parent distribution or the bootstrap method to determine the tail probability points, the data are rank-ordered and the top 100 values are selected; that is, the data in the upper 90th percentile are used to describe the behavior of the tail. An exponential distribution for this tail is assumed, and the parameters of this distribution are determined using the maximum likelihood estimation method. The equation fitted to the original data is

$$F(x) = 1 - \exp[-1.5508(x - 4.1407)]. \quad (12)$$

The fit of the above equation to the data, shown in Fig. 1(b), is quite reasonable. The goodness-of-fit evaluated with the chi-square test suggests that the validity of the above exponential distribution cannot be rejected. The first order statistic based on the exact theory of extreme value is

$$G_1, n = [F(x)]^n. \quad (13)$$

The distribution of the highest values, developed using Equation (12) in (13) for $n = 100$, is presented in Fig. 2(a). It is observed that the maximum value, 8.1, in the original sample occurs at a probability of 80%. That is, there is a 20% probability that this value can be exceeded if additional data are acquired. From the highest value probability distribution, it is seen that there is a 95% probability that the maximum value could lie between 6.3 and 9.5 (see Fig. 2a). For $n = 1000$ and the exponential fit in Equation (13), it is observed that the maximum value, 8.1, in the original sample occurs at a probability of 12%, and the 95% confidence interval for the maximum ranges from 7.75 to 10.97. Using Equation (1) with $m = 1$ and $n = 1000$ and the normal distribution (Equation (6) for $F(x)$ with a mean of 3.4 and a standard deviation of 1.63) yields a range of 7.77–9.95 as the 95% confidence interval for the maximum value. This difference in the upper bound values between the two methods is probably due to the slight overestimation of the maximum by the exponential fit to the tail of the normal distribution (see Fig. 1b). It should be recalled that the confidence intervals estimated from Equation (1) represent the range of outcomes from the order statistics whereas the confidence intervals presented in the first two columns of Table 1 represent the confidence limits to the mean value of the quantile. Although the confidence bounds derived from the direct application of the bootstrap method (column 4 of Table 1) should be comparable to the above confidence bounds from the exact theory, the 95% confidence intervals for the maximum value cannot be compared since the bootstrap method is not reliable to develop confidence intervals for the maximum value. However, at other

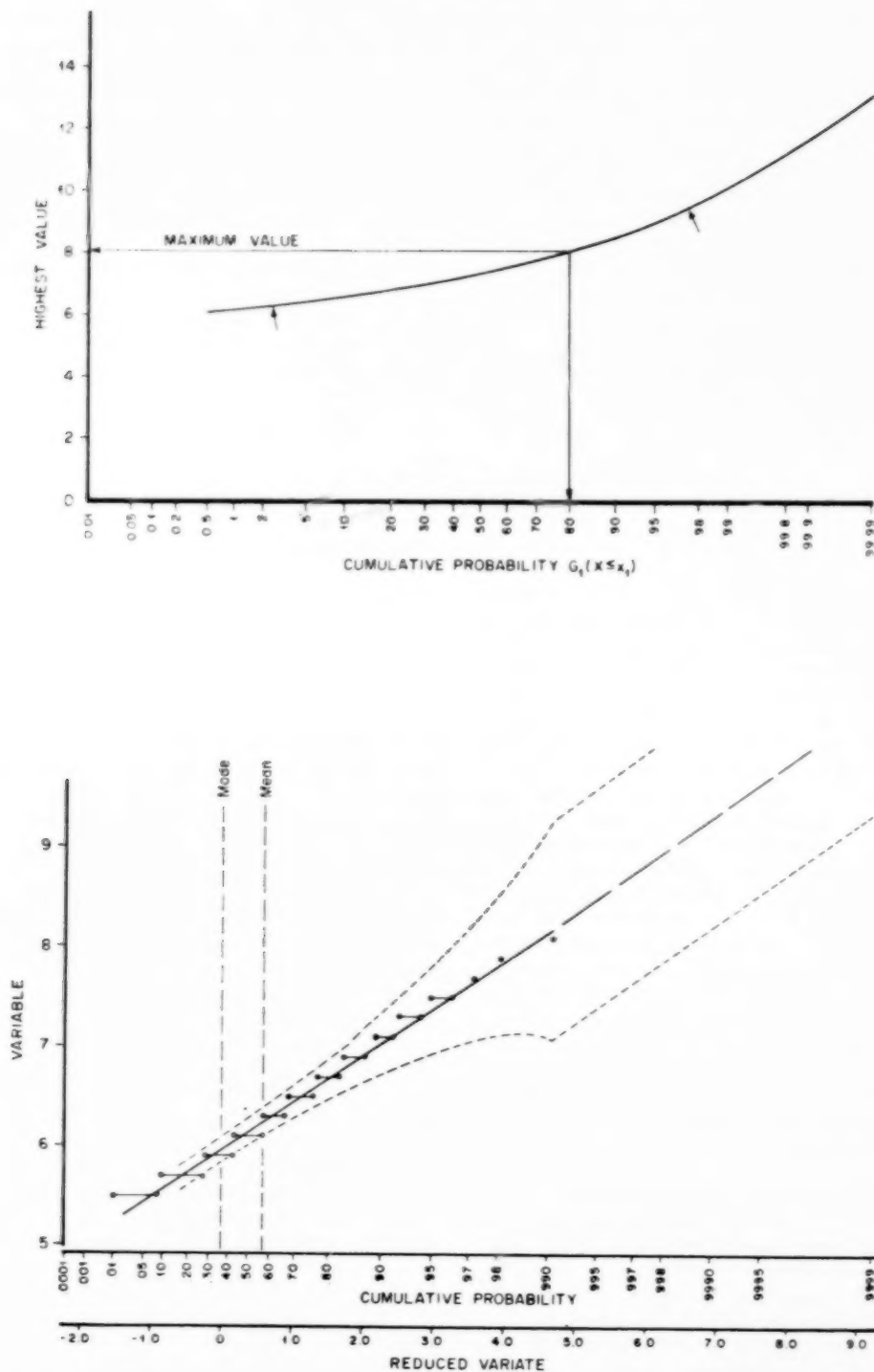


Fig. 2. (a) The highest value distribution for the data from the normal distribution. The arrows indicate the 2.5 and 97.5% probability points from which 95% confidence interval for the highest value can be determined. (b) Asymptotic extreme value distribution for the top 100 values of the data from the normal distribution. These values are now considered to represent the maxima of the variable. The 95% confidence intervals along with the locations for the mean and mode of the distribution are also presented.

quantiles where there are sufficient numbers of samples from which a meaningful resampling can be made, the 95% confidence intervals developed using Equation (1) are comparable to those determined from the direct application of the bootstrap method.

4.1.5. Analysis with the asymptotic theory for extreme values. The top 100 values, used in determining the exponential distribution above, are now assumed to represent the maxima of the random variable. These values are then subjected to analysis using the asymptotic theory for extreme values. The asymptotic distribution is presented in Fig. 2(b) along with the 95% confidence limits to the best-fit line $Y = A(x - U)$. The mean and mode of the extreme value distribution are also indicated in the diagram. From this distribution the highest value, corresponding to a probability of 99.99%, is 8.18 with the 95% confidence limit of ± 1.09 for the maximum. This confidence interval of ± 1.09 for the maximum value is comparable to that obtained from the use of the order statistics based on the full parent distribution in Table 1.

In order to present another method for the bootstrap modification, the sample of 100 values (representing the maxima) are subjected to the bootstrap resampling procedure. In each bootstrap replication, the Type I asymptotic extreme value distribution for the sample consisting of 100 new values chosen randomly with replacement from the original sample is determined. Based on the mode and the Gumbel slope for the bootstrap sample in each replication, the numerical value for the variable corresponding to the 99.99% point (at $Y = 9.21$) is computed. Utilizing 1000 bootstrap replications, the 1000 numerical values corresponding to the 99.99% probability point are determined. From the bootstrap distribution of these 1000 values, it is determined that the numerical values range from 9.1 to 11.5 at this extreme probability point. The extrapolated 95% confidence interval for the best-fit line of the extreme value distribution describing the original sample (see Fig. 2b) indicates lower and upper limits of 9.3 and 11.5, respectively for the variable ($x = 10.4$ at $P = 0.9999$) at the probability of 99.99%. These results suggest that this type of parametric smoothing for the bootstrap technique will enable removal of the inherent constraint with the bootstrap method (this is, values outside those present in the original sample cannot be generated).

The major drawback of using the full parent distribution to properly describe the tails is that one needs to know *a priori* what the true parent distribution is. As stated before, the air pollutant concentrations cannot be described by a single universal distribution that would be valid for all times and locations. Further, one could obtain a good fit of a theoretical distribution in the central portion of the data, but fail to achieve such a fit in the tails. Furthermore, as shown by Hirtzel (1984), the estimates of extremes based on the full parent distribution can differ considerably depending upon the form of the underlying distribution assumed. On the other hand, for the tails of a large number of initial

distributions belonging to the exponential family, either the maxima can be fitted to the Type I asymptotic extreme value distribution, or the tail of the parent distribution can be fitted to an exponential distribution from which the statistical properties of the extreme values can be determined. Thus, the analyses presented here indicate that the bootstrap method for the data in the central portion of the cumulative distribution and the order statistics for the tails of the cumulative distribution enable the estimation of confidence intervals for the entire distribution.

4.2. RAPS data analysis with the extreme value statistics

4.2.1. Application of the exact theory of extreme values. The predicted and measured hourly concentrations at site 108 under each of the 12 meteorological situations over the entire 1-y record are rank-ordered, and the upper 10% of RAM model predictions as well as measured values are fitted to the exponential distributions of the form

$$F(x) = 1 - e^{-\lambda(x-B)} \quad (14)$$

in order to perform extreme value analysis using the exact theory. The method of maximum likelihood estimation for curve-fitting is employed to estimate the parameters λ and B of the distribution. The CDF for the first and second maxima from the hourly values of concentrations under each meteorological situation (sample size n) based on the exact theory are given as

$$G_{1,n} = [P(x \leq x_1)]^n = [F(x)]^n \quad (15)$$

$$\text{and } G_{2,n} = n[F(x)]^{n-1} - (n-1)[F(x)]^n, \quad (16)$$

where $P(x \leq x_1)$ is the probability of any value of x less than or equal to the maximum x_1 , observed over the n values. Since the data are assembled for similar atmospheric conditions, the data set can be thought of as satisfying the identical distribution assumption. Auto-correlation analysis on these data confirms that the data are statistically independent. The effect of autocorrelation on the estimation of the maxima was discussed by Hirtzel *et al.* (1982). The problems caused by the lack of independence of the data appear to be limited to the associated reduction in the number of independent values (Tabony, 1983).

Since the major thrust of this paper is on demonstrating the utility of these statistical techniques in model evaluation, the analytical exponential distributions for both the model predictions and the measured data in the upper 10% of the data under daytime neutral atmospheric conditions only for one outer city site 108 (sample size of 104 measured and predicted values at this percentile) are presented in Fig. 3(a). The actual data along with the best-fit curve for the measured concentrations are indicated in the diagram to provide a visual examination of the fit of the theoretical curve to the actual data. It is seen that the model underpredicts the concentrations for the

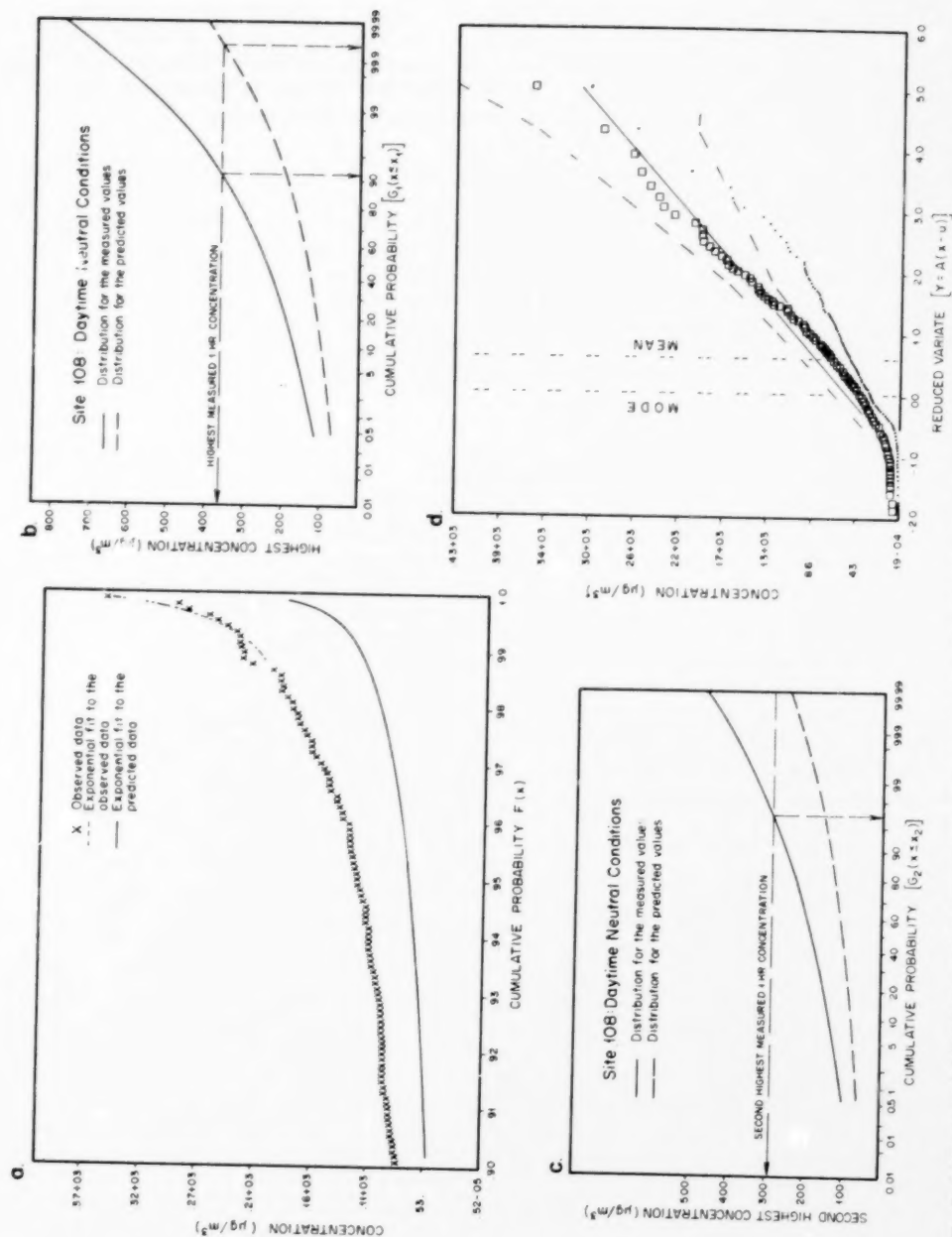


Fig. 3. (a) Exponential distributions for the upper 10% of measured and predicted hourly concentrations under daytime neutral stability at site 108. (b) The highest value distribution for the measured and predicted hourly concentrations. (c) The second highest value distribution for the measured daily maximum 1-h concentrations (squares) under daytime neutral stability condition at site 108. The 95% confidence intervals for the distribution are also presented. The distribution of predicted values (stars) are indicated in the diagram to provide a comparison between measured and predicted extreme values.

entire upper 10% under these neutral atmospheric conditions. Similar plots (not presented here to conserve space) are developed for the data under nighttime neutral, and unstable atmospheric conditions. As before, these plots reveal that the model underpredicts the extreme values. For the data under stable atmospheric conditions, the exponential fit to the upper 10% of the data cannot be rejected. The model overpredicts the concentrations under stable conditions. In order to disclose the model performance under different wind speed conditions, data are segregated according to low ($< 2 \text{ ms}^{-1}$), medium ($2\text{--}5 \text{ ms}^{-1}$), and high ($> 5 \text{ ms}^{-1}$) wind speed conditions, and the upper 10% of the data in each category are subjected to the exact theory of extreme values. It is observed that the model has a tendency to overpredict (underpredict) the concentrations under low (high) wind speeds. Reasonably good agreement is seen between model predicted and observed extreme values under medium wind speed conditions (diagrams for different wind speeds are not presented here). The plausible explanations for this behavior of RAM were discussed in Rao *et al.* (1985) and in Turner and Irwin (1984).

The cumulative probability distributions for the highest measured and predicted 1-h concentration values under daytime neutral conditions at site 108, based on (14) and (15) with the appropriate values for the parameters λ and B , are presented in Fig. 3(b). Models that reasonably simulate the process which produces the measured highest value distribution should have a $\text{PROB}[G_1(x) \leq \text{highest predicted concentration}]$ comparable to that of the measured highest value distribution function (see Rao and Visalli, 1981). Under the daytime neutral conditions at site 108, the highest measured hourly concentration is $354 \mu\text{g m}^{-3}$ and the probability at which this value occurs is 0.91. At the concentration level, the $\text{PROB}[G_1(x) \leq 354 \mu\text{g m}^{-3}]$ is 0.9996 for the RAM model (see Fig. 3b). In other words, whereas the highest value distribution for the measured data indicates that there is 9% chance of a greater concentration occurring than the actual highest value observed, the model indicates that there is only 0.04% chance that this concentration level can be exceeded. The cumulative probability distributions for the measured and predicted second highest hourly concentration under daytime neutral conditions at site 108 based on (14) and (16) are presented in Fig. 3(c). The second highest 1-h measured concentration is $288 \mu\text{g m}^{-3}$ occurring at a probability of 0.97. The model indicates that there is no chance of predicting this level of concentration (see Fig. 3c). If there are relatively small differences between rank 1 values and the remaining values, $\text{PROB}[G_1(x) \leq \text{maximum value}]$ will be low. If $\text{PROB}[G_1(x) \leq \text{maximum value}]$ is extremely high, it may indicate that the maximum value represents a spurious or unusual observed or predicted concentration. On the other hand, if the probability is extremely low it may not be a worthwhile comparison since there would be a great probability that a much higher concen-

tration than currently existed could be measured or predicted.

The cumulative probability distributions for the measured and predicted third, fourth, . . . , m th maximum can also be developed from Equation (1), and the two distributions can be compared. The 95% confidence interval for these extreme concentrations of interest can be determined from the cumulative probability distribution of the first, second, third, . . . , m th highest values; whether the corresponding modeled extreme concentration lies within the confidence interval for the measured extreme concentration can then be determined to quantify the performance of the model in simulating the extreme values of interest. Based on the analysis performed here using the exact theory of extreme values for the upper 10% of the concentrations data at site 108, it is evident that the model overpredicts under low wind speed ($< 2 \text{ ms}^{-1}$) as well as stable (P-G stability classes E-G) atmospheric conditions, and underpredicts for high wind speed ($> 5 \text{ ms}^{-1}$) as well as neutral and unstable atmospheric situations.

4.2.2. Application of the asymptotic theory of extreme values. Hourly concentration data at site 108 for each day are screened to determine the daily maximum 1-h concentration under each meteorological situation. The measured and the predicted daily maximum 1-h concentrations over the entire 1-y record are then segregated according to the atmospheric stability conditions, and subjected to analysis using the asymptotic theory for extreme values. Whereas the data analyzed with the exact theory consists of measured and predicted hourly concentrations in the 90th percentile, the daily maxima of 1-h concentrations are used in the application of the asymptotic theory. In both cases, data under daytime neutral atmospheric conditions only are analyzed here.

The asymptotic extreme value distribution for the measured data under daytime neutral atmospheric conditions at site 108 (sample size of 150 daily 1-h concentration maxima) is presented in Fig. 3(d). The method of maximum likelihood is used to determine the best-fit-line through the measured values (squares) along with the 95% confidence interval for the distribution is indicated in the diagram; the individual model predictions (stars) under similar atmospheric conditions are also presented. One can test for equality of two distributions using, for example, two-sample goodness-of-fit tests. Visual examination reveals that most of the model predicted values lie outside the 95% confidence interval of the distribution of the measured values, suggesting that there are no grounds for comparison between model predictions and measurements; the two distributions are completely different from each other. The distribution of model predictions appears to be different from the distribution of the measured values up to the 97th percentile. Starting with the 97% probability point, the predicted values are within the confidence interval of the measured distribution. It is also evident that the predicted data

below the 94% probability point belong to an extreme value distribution with different parameters; a jump or discontinuity is seen at the 94% point (corresponding to a Y value of 2.78). Under stable and unstable conditions, all the modeled values lie outside the confidence interval for the distribution of measured values (diagrams not presented here). Thus, at the outer city site 108, the model is seen to overpredict (underpredict) under stable (unstable) conditions. However, an agreement between the modeled and measured daily maxima in the upper 3% of the data is observed under neutral atmospheric conditions.

The parameters for the extreme value distribution, as well as the highest measured and predicted daily maximum 1-h concentrations under each atmospheric stability condition at site 108 are presented in Table 2. One could develop the asymptotic extreme value distribution for the daily second maximum 1-h concentrations using Equation (4) to determine the model's ability to predict the daily second maxima of 1-h concentration. The 95% confidence interval offset for the highest and the second highest measured value is obtained, and whether the modeled extreme value is within the confidence bounds of the measured extreme value can be determined, to quantitatively assess the performance of the model. Table 3 presents the central tendency parameter values for various stability regimes. It is evident that there is poor agreement between the modeled and measured means and medians at site 108. Even though the data set is comprised of daily maximum concentrations under each stability situation, these results suggest that the model's performance in simulating the central tendencies of the extreme value distribution at this site is poor.

4.3. RAPS data analysis with the resampling technique

The mean and median (50% probability point) of all the measured daily maximum 1-h concentrations (sample size of 150 values) under daytime neutral atmospheric conditions at site 108 are determined. The cumulative frequency distributions for the mean and the median of the measured data under daytime neutral situations at site 108 are developed using the bootstrap method. In each bootstrap replication, the individual values are chosen randomly and with replacement to develop a bootstrap sample. Figure 4(a) presents the cumulative distribution for the measured means and medians based on 1000 bootstrap replications. The non-parametric standard deviation for the mean or the median can be readily determined from the plot in Fig. 4(a) as

$$\sigma_{\text{boot}} = (T_{0.84} - T_{0.16})/2,$$

where T is the value of either the mean or the median corresponding to the 84% and 16% probability points. Whether the predicted mean and median lie within the 68% interval of the measured mean and median can then be determined to assess the model performance.

The distribution of the differences, under neutral atmospheric conditions during the day at site 108,

$$dT^* = [(T_o - T_p) - (T_o^* - T_p^*)],$$

where T_o and T_p are the mean or the median of the observed and predicted daily maximum 1-h concentrations, respectively, and T_o^* and T_p^* are the corresponding values in each bootstrap replication, is presented as a histogram plot in Fig. 4(b). The distribution of the differences for the median is also shown in Fig. 4(b). Whereas the differences between the measured and predicted means appear to be normally

Table 2. Parameters of the asymptotic extreme value distribution for data under different atmospheric stability conditions at site 108. The corresponding P-G classification for each stability situation is identified in the parentheses

Data subset	Gumbel slope		Mode		Highest value ($\mu\text{g m}^{-3}$)	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
Unstable (A-C)	58.124	30.391	46.209	20.594	329	176
Neutral day (DD)	52.816	38.937	45.526	27.995	354	300
Neutral night (DN)	51.530	41.895	52.206	35.675	270	338
Stable (E-G)	56.789	92.096	45.334	91.834	421	580

Table 3. Comparison between measured and predicted means and medians of the daily maximum one hour concentration under each stability situation at site 108. Results from the asymptotic extreme value distribution are also included in this table

Data subset	Mean ($\mu\text{g m}^{-3}$)				Median ($\mu\text{g m}^{-3}$)			
	Based on the original sample		Based on asymptotic extreme value distribution		Based on the original sample		Based on asymptotic extreme value distribution	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
Unstable (A-C)	79	38	80	38	56	31	68	32
Neutral day (DD)	75	50	76	50	53	41	65	42
Neutral night (DN)	81	57	82	60	66	47	71	51
Stable (E-G)	77	144	78	145	60	120	66	126

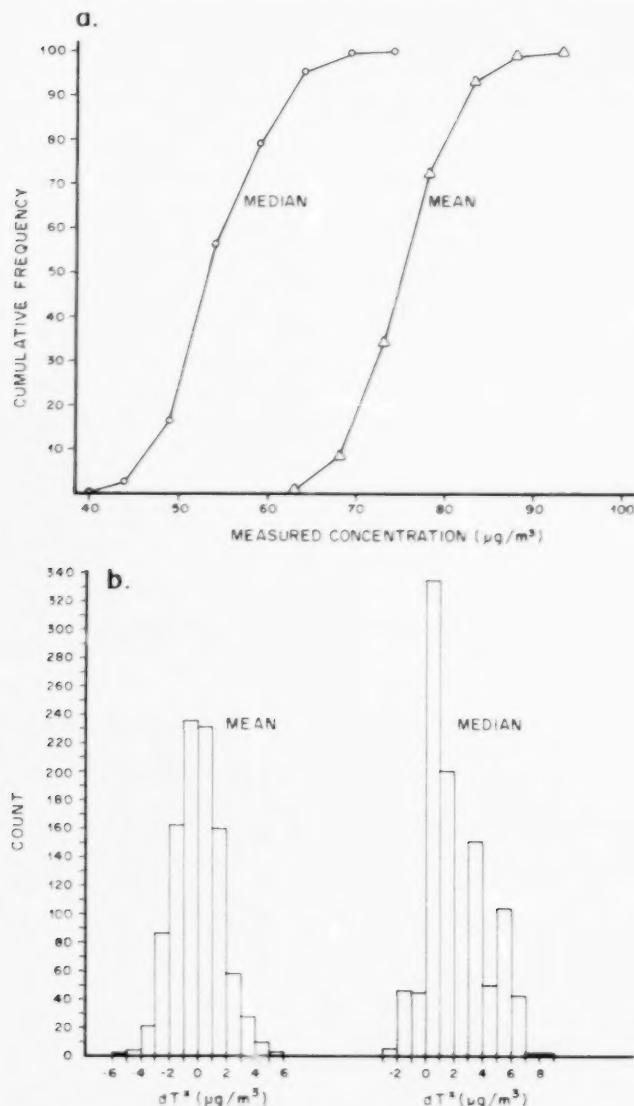


Fig. 4. (a) The cumulative frequency distribution of the means and medians based on the application of the bootstrap technique to the measured daily maximum 1-h concentrations under daytime neutral conditions at site 108. (b) Histogram plot of the differences between measured and predicted means and medians based on the bootstrap analysis of the daily maximum 1-h concentrations under daytime neutral case at site 108.

distributed, the distribution for the median is seen to be non-symmetric. These diagrams represent the uncertainty distributions for the central tendencies of the pollutant maxima. One could calculate σ_{boot} to determine the confidence interval for the differences $F(d)$ based on these distributions. The interpretation of $F(d)$ was discussed by Rao *et al* (1985).

4.3.1. *Determination of confidence intervals.* Although the application of extreme value statistics to the daily maximum 1-h concentrations reveals that the model predictions are within the confidence bounds

for the measured data at the extreme probability points under daytime neutral atmospheric conditions, the use of bootstrap method should enable the quantitative assessment of the model performance without requiring any assumptions regarding the nature of the underlying distribution. The measured daily maximum 1-h concentrations at site 108 under daytime neutral atmospheric conditions over the entire record are subjected to 1000 bootstrap replications. In each replication, the bootstrap sample is rank-ordered, and the value corresponding to each probability point is

determined. The mean of the values at each probability point over the 1000 bootstrap replications is then computed. The approximate 95% confidence intervals at each probability point are determined using the percentile method outlined above. The cumulative distribution for the measured data along with the 95% confidence intervals for the distribution are presented in Fig. 5. The model predictions under similar atmospheric conditions are rank-ordered and the predicted values corresponding to each probability point are shown in Fig. 5. From this diagram, it is evident that most of the model predicted values lie outside the 95% confidence interval for the measured values. It is observed that there is agreement between measurements and predictions beyond the 97% probability point, consistent with the result found with the asymptotic extreme value distribution. Because of the smaller sample size available for resampling at the lower and upper tails, the confidence intervals based on the percentile method to the bootstrap distribution may not be reliable at these probability points. Thus, one cannot assume with the bootstrap analysis that there is in fact an agreement between model predictions and measurements at these extreme probability points. However, in the probability range of 0.20–0.80

of the cumulative distribution for the measurements, where sufficient samples are available for performing the resampling analysis, the predicted values are clearly outside the 95% confidence interval of the distribution of measured data.

In the above analyses, a 95% confidence interval for the observed concentration is chosen and whether the model predicted values lie within this interval is determined. From the results presented in Fig. 5, it is evident that if a different confidence interval were chosen, say 99% confidence interval, this interval would have included most of the model predicted values. *The confidence interval or the performance criteria perhaps should reflect the current level of understanding of the atmospheric processes that affect the transport and diffusion of material in the urban environment.* Any model prediction represents an ensemble average while any given observation reflects a specific realization from a population that will almost always differ from the prediction even if the model and the input data were perfect. In a study to characterize the natural variability of pollutant concentrations due to the stochastic nature of the atmosphere, Eskridge *et al.* (1984) applied the bootstrap method to the concentrations normalized by the emission strength and the

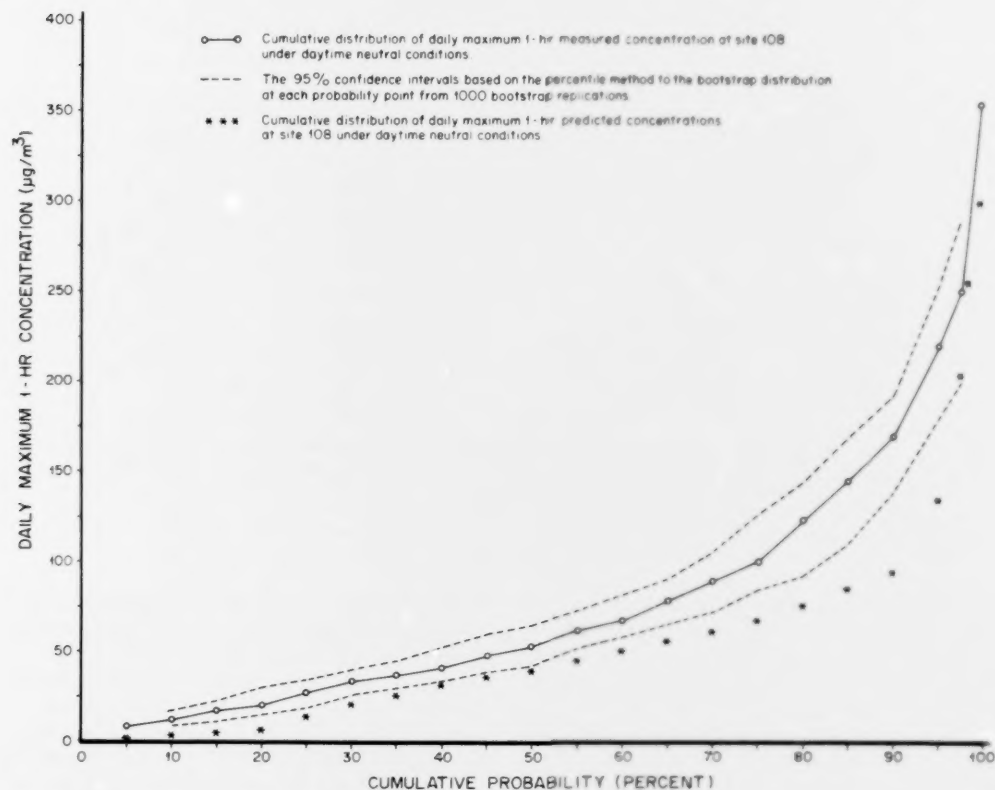


Fig. 5. The cumulative probability distribution of the measured (open circles) and predicted (stars) daily maximum 1-h concentrations under daytime neutral conditions at site 108. The 95% confidence interval for the distribution of the measured values is determined from the bootstrap method.

wind speed, and found that the concentration in a single realization or an event exhibits at least 30% variability or fluctuation from the ensemble mean for a given atmospheric condition. Thus, there is no significant difference between model predictions and observations as long as the predictions are contained within the natural variability of the observed concentrations. In order to reveal the behavior of RAM at site 108, the percent of model estimates of daily maximum 1-h concentration within and outside the range of 30% of the corresponding observed concentrations is presented in Table 4. A significant portion of the model estimates in the underprediction category for neutral and unstable conditions, and in the overprediction category for stable atmospheric conditions is evident. It is apparent that only about 20% of the model estimates are within the range of the so-called natural variability of measured concentrations independent of atmospheric stability.

5. CONCLUSIONS

In this paper, it is observed through a comparison of analyses on a theoretical distribution and extreme value analyses of its tail distributions that the confidence intervals for the values corresponding to the extreme probability points determined from extreme value statistics and bootstrap method will be different from those determined using the full parent distribution and the modified bootstrap method. The reason for this difference is that the percentiles or quantiles estimated by the parent distribution and the modified bootstrap methods involve all of the information about the distribution while the extreme value statistics and the direct application of bootstrap technique do not. It is also observed that the confidence intervals determined from the order statistics and the bootstrap method are in excellent agreement when the underlying distribution for the tail probability points is known *a priori*.

It is demonstrated that the exact and asymptotic theories of extreme values and the bootstrap technique are useful tools for the quantitative evaluation of air quality model performance. The concentration data at site 108 in the RAPS network, divided into four stability and three wind speed classes, are analyzed

by these techniques. It is found that both the measured and predicted data in the upper percentiles can be described by an exponential distribution. The results obtained at site 108 using the extreme value statistics show that the model overpredicts under low wind speeds ($< 2 \text{ m s}^{-1}$) and stable atmospheric conditions, and underpredicts for high wind speeds ($> 5 \text{ m s}^{-1}$) and neutral and unstable situations. Resampling analysis is used to quantify concomitant uncertainty of the model predictions.

The approach utilized here is to develop the cumulative distribution of measured concentrations under specific atmospheric conditions and its 95% confidence intervals based on parametric or non-parametric methods; the cumulative distribution of model predicted concentrations under similar atmospheric conditions are then simply superimposed on the distribution of the measured concentrations (assuming that the measurements are error-free) to determine if the model predictions lie within the confidence bounds for the measured values at each probability point. Only the confidence intervals for the distribution of the measured concentrations in a given meteorological situation are considered here since the atmospheric phenomena that affect the pollutant concentrations are stochastic. This could result in a percentile value from the measured concentrations having a greater variability than that of the modeled value because the air quality models are deterministic. *The variability or uncertainty in the model predicted value at a given percentile stems from the inaccuracies in the input data to the model as well as the physics included in the model.* Because of the inherent limitations to the bootstrap method at the lower and upper tails of the concentration distribution, the extreme value theory for the tails is needed. Thus, with the use of the bootstrap technique and the discriminating capability of the 'tail fit' type model, it is possible to evaluate the air quality model performance over the entire range of the concentrations' cumulative distribution.

Acknowledgements—One of the authors (S. T. Rao) gratefully acknowledges the support of this investigation by the U.S. Environmental Protection Agency under Grant No. CR810646-01. The authors thank William Cox and Cynthia Hirtzel for providing critical review and comments on the manuscript. Thanks are extended to Mr. Garry Lanphear and

Table 4. Percentage of model estimates of daily maximum 1-h concentrations within and outside the range of 30% of their corresponding observed concentrations for each stability regime

Atmospheric stability	% of predicted concentrations within $\pm 30\%$ of the corresponding observed concentrations	% of predicted concentrations $> 30\%$ of the corresponding observed concentrations	% of predicted concentrations $< 30\%$ of the corresponding observed concentrations
Unstable (ABC)	20	10	70
Daytime neutral (DD)	21	22	57
Night-time neutral (DN)	20	23	57
Stable (E-G)	20	59	21

Mrs. Carol Clas for the drafting of the diagrams, and to Ms. Betty Cataldo for typing the manuscript.

REFERENCES

- Breiman L., Stone C. J. and Gins J. D. (1979) New methods for estimating tail probabilities and extreme value distributions. TSC-PD-A226-1, Technology Service Corporation.
- Breiman L., Stone C. J. and Gins J. D. (1981) Further developments of new methods for estimating tail probabilities and extreme value distributions. TSC-PD-A243-1, Technology Service Corporations.
- David H. A. (1981) *Order Statistics*. John Wiley, New York.
- Efron B. (1982) The Jackknife, the Bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Eskridge R. E., Petersen W. B., Rao S. T. and Sistla G. (1984) Turbulent diffusion behind vehicles: experiments and verification of roadway models. *Proc. 77th Annual APCA Meeting*, San Francisco, CA.
- Fox D. G. (1981) Judging air quality model performance. *Bull. Am. met. Soc.* **62**, 218-221.
- Fox D. G. (1984) Uncertainty in air quality modeling. *Bull. Am. met. Soc.* **65**, 27-36.
- Galfond G. J. (1982) Robust estimation of extreme quantities. Tech. Rep. No. 60, Dept. of Statistics, Stanford University, Stanford, CA.
- Gumbel E. J. (1962) Statistical theory of extreme values (main results). In *Contributions to Order Statistics* (edited by Sarhan A. E. and Greenberg B. G.), pp. 56-93. Wiley, New York.
- Hirtzel C. S., Corotis R. B. and Quon J. E. (1982) Estimating the maximum value of autocorrelated air quality measurements. *Atmospheric Environment* **16**, 2603-2608.
- Hirtzel C. S. (1984) Analysis of extreme values of natural processes: statistical description of the maximum. *Appl. Math. Comput.* **15**, 283-303.
- Rao S. T. and Visalli J. R. (1981) On the comparative assessment of the performance of air quality models. *J. Air Pollut. Control Ass.* **31**, 851-860.
- Rao S. T., Sistla G., Pagnotti V., Petersen W., Irwin J. and Turner B. (1985) Evaluation of the performance of RAM with the Regional Air Pollution Study Data Base. *Atmospheric Environment* **19**, 229-245.
- Schiermeier F. A. (1978) Air monitoring milestones: RAPS field measurements are in. *Envir. Sci. Technol.* **12**, 6.
- Tabony R. C. (1983) Extreme value analysis in meteorology. *Met. Mag.* **112**, 78-98.
- Turner D. B. and Novak J. H. (1978) User's Guide for RAM. EPA-600/8-78-016 a and b, U.S. Environmental Protection Agency.
- Turner D. B. and Irwin J. S. (1984) The relation of urban model performance to stability. Preprints of 14th International Technical Meeting on Air Pollution Modeling and Its Application, NATO/CCMS, Copenhagen, Denmark.

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.